Risk Analysis for Resource Planning Optimization

Kar-Ming Cheung*
Jet Propulsion Laboratory, Pasadena, USA

Abstract: This paper describes a systems engineering approach to resource planning by integrating mathematical modeling and constrained optimization, empirical simulation, and theoretical analysis techniques to generate an optimal task plan in the presence of uncertainties. We assume one has a collection of tasks, each having either a known duration or a random duration defined by a probability distribution, and each requiring certain resources (e.g. workforce, budget, etc.) during execution. There can be various constraints operating among the system of tasks; for example, one particular task might be paired for execution with another task, and the first one might be required to be completed before the second can begin. In this work we also assume, just as would be the case in any realistic scenario, that limits are set for resources consumed per unit of operating time, and so there cannot be too many tasks going on at once. The overall goal of the planning study is considered successfully met when one has found an event schedule that is at least a very good local optimum, with no (or only minimal) violations of the constraints.

We recently demonstrated a constraint modeling and optimization framework to support spacecraft non-deterministic event planning and sequencing [1]. We extend the results in [1] for general resource planning problems and introduce a number of improvements.

Keywords: *risk analysis, resource planning, optimization*

1. INTRODUCTION

In [1], we demonstrated a risk analysis approach in the area of non-deterministic event schedule planning and sequencing. In this paper, we improve upon this technique and apply it to the boarder class of resource planning and optimization problems.

Using the problem formulation as discussed in [1], a plan consists of a number of tasks. By planning, we mean the process of a-priori scheduling the tasks with knowledge of the planning horizon and its influences. When the tasks are executed according to plan, the tasks might consume one or more bounded resources that are either replenishable or non-replenishable. We further assume that there is a set of rules and constraints governing the required relationships and dependencies among the tasks. A plan is defined to be successfully executed if 1) all tasks can be accommodated within the planning horizon (if specified[1]); 2) there is no resource usage that exceeds the maximum allowable limit; and 3) there are no violations to the set of pre-defined rules and constraints. The challenge is to determine a conflict-free plan that optimizes a figure of merit (FOM) scoring the plan's quality.

Much work has been done in the area of constrained optimization algorithms to support resource planning applications in which the optimization problem consists of tasks with deterministic resource usage and time durations. Some of these algorithms have been internally implemented in commercial-off-the-shelf (COTS) planning and scheduling tools (ILOG's *ILOG Scheduler* [2] is an example), and a wide variety are available as stand-alone modules for solving general optimization problems. The details of these algorithms is not the emphasis of this paper, although it is worth

**Kar-Ming.Cheung@jpl.nasa.gov*

[1] Not all problems require a finite planning horizon, but rather have the aim simply to complete as many tasks as early as possible. The examples described in this paper mostly belong to this type, though we have solidified the mathematical modeling of the former type as well.

mentioning that some algorithms are a good match to certain types of problems, while others are not; an analyst must always be cognizant of any algorithm's strengths and weaknesses.

In our research, we have found the Matlab [3] platform an excellent tool, and we have extensively used its Optimization Toolbox algorithm *fmincon*, a modern sequential quadratic programming implementation, for generating interim solutions and for various systematic investigations. This is not to say that this algorithm is perfectly ideal, for an algorithm that can solve all constrained optimization problems without difficulty does not exist. Conceptually, with computational resources available, one can apply multiple constrained optimization algorithms to a given problem to ensure that at least one algorithm would converge and deliver a good plan.

In the presence of a planning horizon, deriving an optimal schedule for a collection of events must take this cutoff time into account, and all constraints must reflect whether or not each involved event has been completed before the planning horizon lapses. The solution of the optimization problem will determine the best event execution plan and schedule for all events.

In this paper we address resource planning problems wherein the planner needs to schedule the resources in advance without complete knowledge of all the factors that influence the resource usage of the plan. A plan usually consists of some tasks in which their resource usage and time durations are known, and other tasks in which their resource usage and time duration are non-deterministic. There is no guarantee that when the plan is executed, the scheduled tasks would not violate any pre-defined rules and constraints, and the resource usages would not exceed their maximum allowable limits. Also resources like workforce and schedule are desirable to be maintained at a steady level during a plan's execution. The idea of scheduling tasks into a conflict-free plan becomes obscure and intangible when task durations and their resource usage are not known in advance. Resource allocation decisions are sometime made based on the "gut feeling" (intuition) of experienced planners. This might result in either not enough planning margins or the use of overly conservative margins, which in turn leads to:

1. Conservative plan in which resources are not efficiently utilized, or
2. Aggressive plan that fails to meet the requirements down the road.

In this paper we investigate and formulate a new model-based engineering resource planning approach by introducing novel risk analysis techniques to resource planning optimization in the presence of uncertainties. This approach is an extension and generalization of the telecommunication link analysis technique, which is a proven statistical estimation technique for evaluating communication system performance and trade-off. The new planning approach assumes that we have the statistical descriptions of resource usage and task durations. This assumption might seem implausible at first, but decades of experience in telecommunication link analysis indicates that imposing reasonable probabilistic models (e.g. worst-case models) on the link parameters that are not well-characterized statistically, the "Law of Large Number" still allows one to reasonably and accurately predict the link margin to guarantee the reliability of the link. We apply similar techniques to the resource planning process to quantify the trade-off between risk and performance, as well as to make forward-looking choices.

This rest of this paper is organized as follows: Section 2 describes the problem in some detail. Section 3 introduces some mathematical models of constraints and good objective functions. Section 4 describes an iterative approach that bounds the risk for resource planning optimization. Section 5 discusses in detail a number of analysis and simulation techniques. Section 6 discusses the concluding remarks.

2. PROBLEM DESCRIPTION

In this paper, we introduce the notion of risk in terms of the probability that the plan fails to execute successfully, which we denote as P_F. The failure can be attributed to not meeting any one or more of the conditions as discussed in Section 1, namely, 1) all tasks can be accommodated within the planning horizon, if specified, 2) there is no resource usage that exceeds the maximum allowable limit, and 3) there is no violation to the set of pre-defined rules and constraints. The goal is to determine an optimal plan that minimizes the resource usage that when executed, has a P_F that is below a given tolerable level.

In any general optimization problem with constraints, there is the possibility of having equality constraints, inequality constraints, or both simultaneously. Further, each constraint type can be linear or nonlinear. Often, problems with equality constraints are easier to solve than those with inequality constraints, but this statement depends completely on the forms of the functions involved.

The method of Lagrange multipliers [4] would apply in the case of equality constraints, but that method may still not lead to a system of equations with a solution in closed form. Further, the method of Lagrange multipliers always depends on one having all function derivatives available, which is often not the case.
When inequality constraints are present, the problem is one involving the Karush-Kuhn-Tucker (KKT) system of several conditions [4], and this is significantly more complicated to apply than the method of Lagrange multipliers, also requires derivatives of all functions to be available, and very rarely leads to accessible solutions.

In most of our studies, and in the topics covered in this paper specifically, modern numerical methods are absolutely essential; the optimization problems encountered are high-dimensional and the theoretical characterizations of optima are not at all practical for finding those solutions.

3. MATHEMATICAL MODELS OF CONSTRAINTS AND OBJECTIVE FUNCTIONS

This section outlines the mathematical models of some objective functions and constraints that are useful for resource planning and scheduling. We use the following notations for this paper: Consider n events of interest and the planning horizon $[T_s, T_e]$, and T_e-$T_s \equiv T$. Let event E_i be characterized by the ordered pair (t_o^i, d_i) for $1 \le i \le N$, where t_o^i and d_i are the start-time and duration of event E_i respectively. Let t_o^i be bounded by $[T_{\min}^i, T_{\max}^i]$, and let d_i have a probability distribution $p_i(d_i)$. For the sake of simplicity, let the event durations d_i's be independent of each other, and let $p_i(d_i)$ have a unimodal probability distribution function characterized by m_i and σ_i, where m_i and σ_i are (respectively) the mean and standard deviation of d_i. To support constrained optimization, we choose fixed event durations $\Delta_i = m_i + \lambda_i\sigma_i$ such that $d_i \le \Delta_i$ with a reasonably high degree of confidence.

3.1 Objective Functions

In seeking an optimal schedule, it is first necessary to define the criterion of optimality/figure of merit (FOM) and set up an objective function to measure the merit of any given choice of event schedule. Often in practice, a "good" schedule is one that accomplishes as much as possible, as soon as possible, subject to whatever demands and constraints are present. This would especially be the case in planning for mission success in a harsh environment or where significantly reducing the probability of undesirable, unplanned outside influences is impossible. From this viewpoint, a good objective function for this optimization problem can be given by

$$f_1(t_0^1,\ldots,t_0^n) = \sum_{i=1}^{n} t_o^i. \qquad (1)$$

The objective function could take on other forms. For example, it can be a weighted sum of the t_o^i values; if we define each event E_i to have a priority weight w_i, another good objective function is given by

$$f_2(t_0^1,\ldots,t_0^n)=\sum_{i=1}^{n} w_i t_o^i \qquad (2)$$

This objective function allows one to adjust the significance of each event's time of initiation independently, and even zero could be used for certain w_i values if the associated t_o^i values were deemed to not be significant drivers of the solution's quality.

3.2 Constraints

If the events are all independent, the problem can be as simple as populating the timeline(s) with as many events as possible. But in most real-world scenarios, many events of interest are dependent on each other in one or more ways. These dependencies can be expressed in the form of rules and constraints. Also, the planning horizon T is usually not long enough to accommodate all events. Therefore, a criterion of optimality is to fit as many high-priority events as possible into the planning horizon without violating any constraints. The dependencies between groups of events can give rise to constraints that are (sometimes) complicated; we will not attempt to describe them in much detail here. The following are some examples of constraints among events. The detailed derivations of these constraints can be found in [1]. For the sake of simplicity, we do not assume a finite planning horizon in the constraints shown here, though in [1] the full forms with a finite planning horizon are given.

Time window: A time window-type of constraint requires that event i falls within a specified time frame $\left[T_{\min}^i, T_{\max}^i\right]$, regardless of whether it occurs within the planning horizon or not. That is $T_{\min}^i \le t_o^i \le T_{\max}^i \quad \forall i$. So if we define $X=[t_o^1,\cdots,t_o^n]^T$ (vector of state variable), $L=[T_{\min}^1,T_{\min}^2,\cdots T_{\min}^n]^T$,and $U=[T_{\max}^1,T_{\max}^2,\cdots T_{\max}^n]^T$, we have the following linear constraints:

$$L \le X \le U \qquad (3)$$

Time order: if E_i and E_j are both scheduled, E_i must start before E_j finishes for some i and j.

$$t_o^i - t_o^j - \Delta_j \le 0 \qquad (4)$$

Inclusion: if E_i is scheduled, then E_j must be initiated in some chosen time interval $[w_o^j, w_f^j]$.

$$\left|2t_o^j - w_o^j - w_f^j\right| + w_o^j - w_f^j \le 0 \qquad (5)$$

Exclusion: if E_i is scheduled, E_j must not be initiated in some chosen time interval $[w_o^j, w_f^j]$.

$$w_f^j - w_o^j - \left|2t_o^j - w_o^j - w_f^j\right| \le 0 \qquad (6)$$

Forbidden Synchronic: if E_i and E_j occur, they must not overlap.

$$\max(t_o^i+\Delta_i, t_0^j+\Delta_j) - \min(t_o^i, t_0^j) \ge \Delta_i + \Delta_j \qquad (7)$$

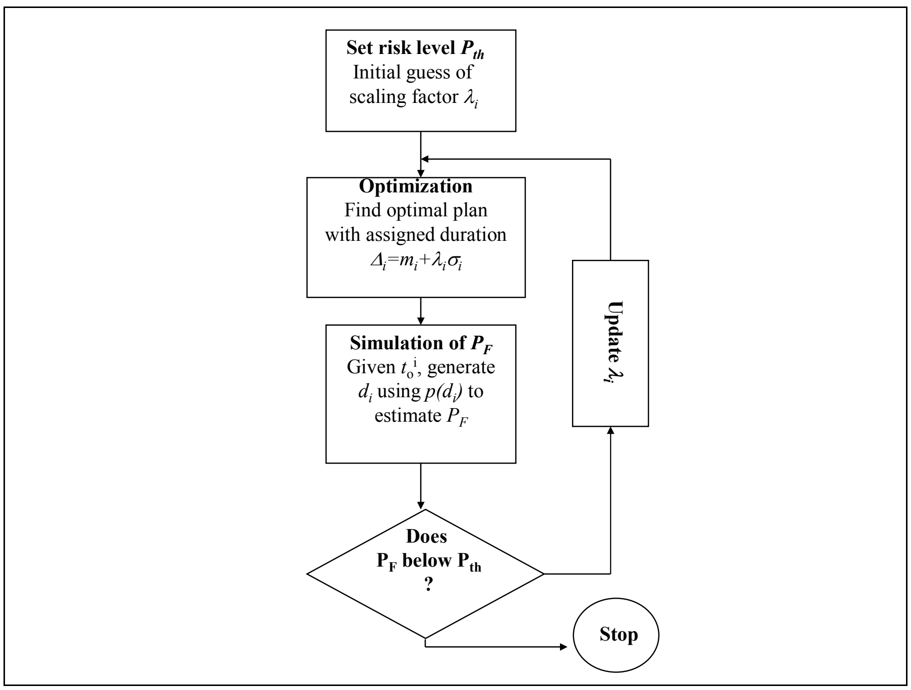
4. ITERATIVE APPROACH FOR RISK-BASED RESOURCE PLANNING OPTIMIZATION

This Section presents an analysis approach that iteratively applies constrained optimization and Monte Carlo simulation to reach an optimal plan that is free of constraint violations with an acceptable P_F. This approach was first suggested by William Gearhart in an internal JPL study titled "Non-Deterministic Sequence Validation and Verification." [5], and later extended to support risk analysis for mission planning in described in [1].

We use the following procedure that iteratively applies constrained optimization and Monte Carlo simulation to reach an optimal plan as illustrated in Figure 1.

- Note that this plan is intentionally "sub-optimal". Once a plan is generated, the start times t_o^i's are fixed, regardless of how the events are executed. That is, the start time t_o^i of event E_i is not dependent upon the completion time of any prior events. This guarantees successful execution of the plan as long as $d_i \le \Delta_i$ for all i.

Figure 1: Iterative Procedure to Find Optimal Plan



We applied the above iterative process to an example of a 10-event case with the following constraints:

- Events 1 and 3 may not overlap
- Event 1 must finish before Event 4 begins,
- There is only one type of resource consumed, and all events consume that resource at a rate of one resource unit per time unit, with the maximum allowable consumption at any time to be 3 units at any time.

Table 1 summarizes the probability distribution and its parameters of each of the 10 event durations.

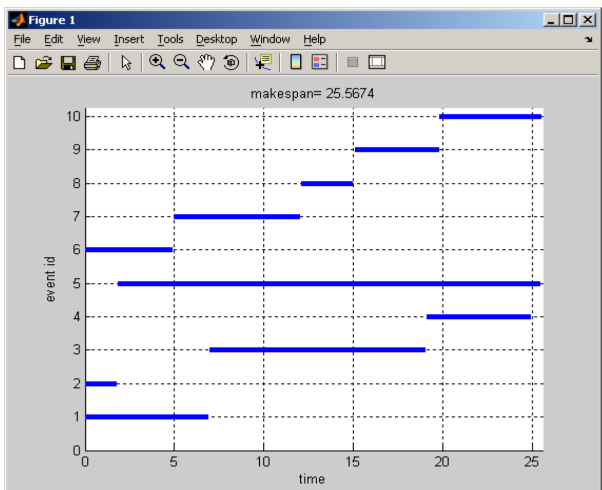
Table 1: Probability Distributions of Event Durations

Event ID	Type of Dist.	Parameters	Min. Value	Max Value

1	Uni.	NA	5	7
2	Beta	α=4, β=4	1	3
3	Norm	μ=10, σ=.5	NA	NA
4	Tri.	Peak=4	3	5
5	LogN	μ=2, σ=.5	NA	NA
6	Uni.	NA	2	5
7	Beta	α=5, β=5	3	8
8	Uni.	NA	1	3
9	Tri.	Peak=3	2	5
10	Tri.	Peak=4	2	6

As in the 5-event case, events in this example are scheduled optimally, with duration of each event fixed at the value where each event has a 99% chance of taking that long or less to complete. Again we apply FMINCON to optimize the above plan of 10 events and Figure 2 illustrates the resulting timeline.

Figure 2: Optimal Plan



Again we perform 10 Monte Carlo simulations of 5000 runs each to estimate P_F . The results are tabulated in Table 2 below.

Table 2: Simulations of P_F

Simulation ID	Probability of Schedule (10 Events) Failing
1	0.0424
2	0.0430
3	0.0458
4	0.0448
5	0.0382
6	0.0372
7	0.0358
8	0.0434
9	0.0400
10	0.0430

Average P_F	0.0414
Upper Bound of P_F	0.10

5. ANALYSIS AND SIMULATION TECHNIQUES

In this section, we discuss a number of analysis and simulation techniques that facilitate the risk analysis for the resource planning optimization process.

5.1 A Simple Upper Bound of P_F

This upper bound was derived in [1], and we simply state the result in this paper. Let $P_{F,i}$ be the probability that task i fails to complete within the pre-assigned time duration, $1 \le i \le n$, an upper bound of P_F given by

$$P_F \le P_{F,1} + P_{F,2} + \cdots P_{F,n} \tag{8}$$

Note that this simple upper bound of P_F does not require tedious Monte Carlo simulations of the plan, and is just the sum of all $P_{F,i,} 1 \le i \le n$. Also this upper bound is independent of the optimization algorithm used to generate the plan.

5.2 Saddle-Point Approximation of P'_F of an Ensemble of Tasks in Tandem

In task planning, a common situation is that there are a number of tasks that are required to execute in tandem. If there are no other dependencies between any tasks in this ensemble with other tasks in the plan, one can merge this ensemble of tasks into one task to estimate the overall task duration and its resource usage. This would simplify the downstream analysis and optimization processes.

In [6], we use a variant of Saddle-Point approximation to estimate the tail probability that the sum of data generated from a number of instruments exceeds a given pre-planned value. This technique can be used to estimate the probability of failure P'_F to execute the series of tasks within the pre-assigned time duration. Let z denotes the ensemble and $x_1, x_2, \cdots, x_n$ denote the durations of n tasks in tandem, where x_i has a pdf $f_{x_i}(x_i)$ for $1 \le i \le n$. That is, $z = \sum_{i=1}^{n} x_i$. Let $f_z(z)$ denote the pdf of z, and $\Psi_z(s) = \int_{-\infty}^{\infty} e^{sz} f_z(z)dz$ denote the characteristic function of $f_z(z)$. The straightforward approach to evaluate $f_z(z)$ as the convolution of $f_{x_1}(x_1), f_{x_2}(x_2), \cdots f_{x_n}(x_n)$ is usually impractical, as this involves $n-1$ integrations. Another approach is to evaluate the characteristic function $\Psi_z(s)$ of $f_z(z)$, which is the product of the characteristic functions $\Psi_{x_1}(s), \Psi_{x_2}(s), \cdots \Psi_{x_n}(s)$ of $f_{x_1}(x_1), f_{x_2}(x_2), \cdots f_{x_n}(x_n)$. The problem with this approach is the difficulty of inverting $\Psi_z(s)$, which can be a very complicated expression, back to $f_z(z)$. Helstrom [7] described a variant of the saddle-point approximation that estimates the tail probability $q_+(\alpha) = \int_{\alpha}^{\infty} f_z(z)dz$. This method is useful in the case where the pdf $f_z(z)$ can be arbitrarily complicated but its characteristic function $\Psi_z(s)$ is known. This approximation is particularly good for small $q_+(\alpha)$. The key result is that the approximation of $q_+(\alpha)$ can be expressed as a function of the characteristic function $\Psi_z(s_o)$ its first derivative $\Psi'_z(s_o)$, and its second derivative $\Psi''_z(s_o)$, where s_o is a positive root of some function $\psi(s)$, and it is shown in [7] that the root exists. There is no need to invert $\Psi_z(s)$. A simpler proof of this approximation can be found in [8], and the main result is described below:

Define the function $\psi(s)$ as follows:

$$e^{\psi(s)} = \frac{e^{-s\alpha}\Psi_z(s)}{s} \tag{9}$$

It was shown [3] that the solution of $\psi'(s) = 0$ exists, and is denoted by s_o. By applying a Taylor series expansion of $\psi(s)$ truncated at the 2nd-order term, it can be shown that the tail probability $q_+(\alpha)$ can be approximated by

$$q_+(\alpha) \approx \frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}} \tag{10}$$

It was shown in [7] that (9) suffices for most engineering applications when α is at least one standard deviation from the mean. $q_+(\alpha)$ is a good approximation for P'_F

5.3 Using Stochastic Optimization Methods to Find a Good Initial Solution

In our research and experiments we have found the Matlab platform an excellent tool, and we have used Matlab's Optimization Toolbox routine *fmincon* as a major computational engine for the non-stochastic solving of general constrained optimization problems. The *fmincon* algorithm implements the Sequential Quadratic Programming (SQP) method [9], finding a minimum of an objective function subject to linear and/or nonlinear constraints, with equality- and inequality-type constraints both enforceable simultaneously. Since *fmincon* is set up natively as a minimization algorithm (optimization theory and practice is fundamentally identical whether one is minimizing or maximizing), the objective function we use in this problem must be designed so that increasingly "bad" task schedules produce increasingly large objective function values. The optimization algorithm will then minimize the "badness" in the schedules, producing good schedules that meet the desired constraints. Note that the SQP method for solving constrained optimization problems can usually reliably find locally-optimal solutions only when all functions involved are continuous, and the starting point for the method (also called the initial guess) is critically important to the quality of the final solution and the number of iterations required to find one. In fact, for a problem of this type, poor initial points will often result in failure of the solver to find *any* feasible solution at all. Conversely, when the solver is started at a point that is close to a good locally-optimal solution, the SQP method can often zoom in to that solution quickly and accurately.

To find a good initial point, we resort to software-based heuristic methods such as genetic algorithms (GA) and/or stochastic algorithms in the Markov Chain Monte Carlo (MCMC) class, including simulated annealing, Gibbs sampling, and direct random sampling from the conditional density kernel of the objective function *given* that the constraints are satisfied. Genetic algorithms and the stochastic MCMC methods do not require functions that are differentiable or even continuous in order to be applied to the problem, and the theory of some MCMC methods (unlike that of SQP) usually gives them the advantage that under rather mild conditions they can converge on very high quality solutions when well-tuned; often even globally-optimal ones.

The GA approach is a set of software-based heuristics that use search techniques inspired by evolutionary biology, such as mutation, selection, gene crossovers, and inheritance. In a genetic algorithm multiple candidates in each subsequent "generation" are chosen stochastically for recombination, using some fitness measure/function. This process simulates the higher probability of survival to reproductive maturity of the organisms (solutions) with higher fitness. These "surviving" candidates are then recombined and randomly mutated to form the next generation(s), with this process going on repeatedly for many, many generations (iterations.) Methods based on these biological processes are not provably MCMC methods and do not have the mathematical rigor of the MCMC algorithms, but some can still potentially be effective if they are well-tuned for the problem. The structural and numerical tuning required for an effective GA is always a case-by-case endeavor and is not easy to encapsulate in few words. An effective genetic algorithm may have a tendency to find good solutions when the search space is large and those solutions are not isolated into regions surrounded by large amounts of infeasible space.

The methods outlined above give rise to a two-step solution procedure that we have applied to these schedule planning problems to achieve very good solutions. By running a hybrid scheme consisting of a stochastic optimization method or GA followed by a non-stochastic algorithm such as SQP, we can typically obtain solutions that are significantly better than either method running alone would find. In one recent experiment on a nontrivial test problem of the aforementioned type with a known unique globally-optimal solution but with non-convex constraints, a stochastic method precursor followed by a run of *fmincon* succeeded in finding that solution.

In the next diagrams we show the outcome of our two-step approach to a resource planning scenario that involves 100 tasks, 40 constraints, and two resources of limit 4. The resulting solution plan is shown in Figure 2, and the resource usage profile for resource 1 is shown in Figure 3. Note that most of the tasks can be accommodated within the planning horizon, and there is no resource usage violation for resource 1[2].

Figure 2: Optimal Plan of 100 Events

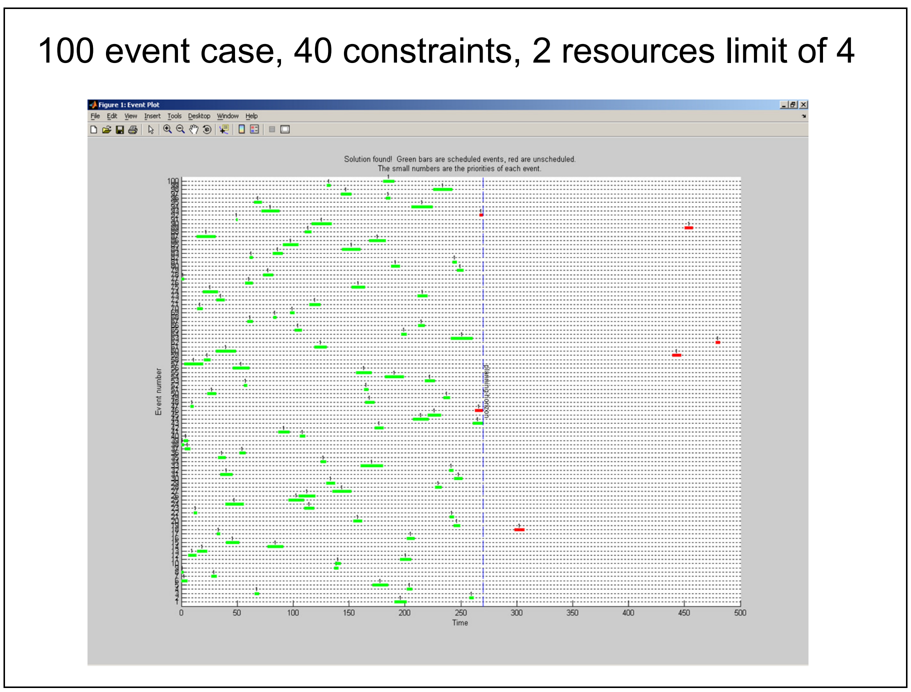
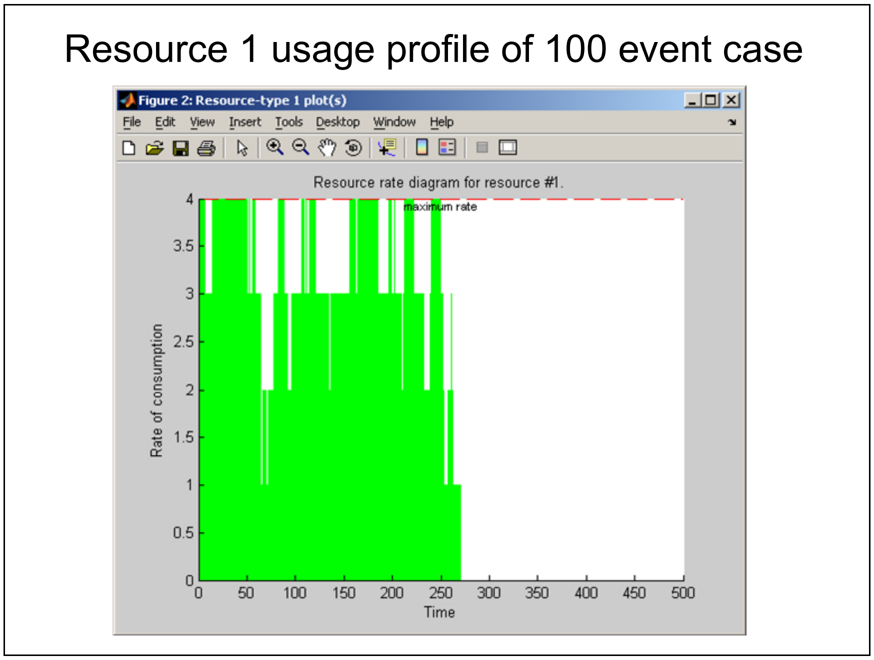


Figure 3: Usage of Resource 1

[2] There is no resource usage violation of resource 2 either, but it is not shown here.



6. CONCLUSION

In this paper, we describe a system engineering approach for resource planning by integrating mathematical optimization, empirical simulation, and theoretical analysis techniques to generate an optimal task plan in the presence of uncertainties. This approach introduces risk analysis techniques to the resource planning process to quantify the trade-off between risk and performance to allow planners and decision makers to make forward-looking choices.

Acknowledgements

The author would like to thanks David Heckman for his insightful discussion on stochastic optimization methods and proofreading the paper. This research was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

References

[1] K. Cheung, A. Ko, V. Dang, D. Heckman, "*Risk Analysis for Non-Deterministic Mission Planning and Sequencing*," IEEE Aerospace Conference, Big Sky, Montana, March 6-12, 2005.
[2] ILOG Scheduler, ILOG Inc., http://www.ilog.com/products/scheduler/
[3] Optimization Toolbox, The Mathworks, http://www.mathworks.com/products/new_products/
[4] Hillier & Lieberman, "*Introduction To Operation Research*," Seventh Edition, McGraw-Hill, 2001.
[5] W. B. Gearhart, "*Non-Deterministic Sequence Validation and Verification*," JPL Internal Report, June 10, 2004.
[6] K. Cheung, "*Probabilistic Instrument Data Generation Analysis Using a Variant of Saddle-Point Approximation*," International Planetary Probe Workshop 5, Bordeaux, France, June 23-29, 2007.
[7] C. W. Helstrom, *Approximate Evaluation of Detection Probabilities in Radar and Optical Communications*" IEEE Trans. Aerosp. Electron. Syst.; vol 14, pp.630-640, July 1978.
[8] M. Emami, S. Zahedi, http://www.stanford.edu/~memami/ee359/final.html
[9] Han, S.P., A Globally Convergent Method for Nonlinear Programming, Journal of Optimization Theory and Applications, Vol. 22, No. 3, pp. 297-309, 1977.